ARTICLE IN PRESS

Computers & Industrial Engineering xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Organizational learning: Approximation of multiple-level learning and forgetting by an aggregated single-level model

Ilhyung Kim[a], Mark Springer[a,*], Zhe George Zhang[a], Young-Sun Park[b]

[a] College of Business and Economics, Western Washington University, Bellingham, USA
[b] Division of Business Administration, Seokyeong University, Seoul, South Korea

ARTICLE INFO

Keywords:
Learning curves
Forgetting
Knowledge management
Multiple-level learning
Liberty ships

ABSTRACT

In a large organization, learning and forgetting may occur at different rates at the various levels of the organization. Recently, it has been shown that a multiple-level learning model works effectively for the accurate measurement and prediction of learning and forgetting in such an organization. Due to a lack of sufficiently detailed data at each organizational level, however, it is often necessary to use the conventional aggregated single-level model to estimate the learning and forgetting of the entire organization. In such an approximation, the potentially different impacts of learning and forgetting at different levels of the organization is not explicitly considered. This paper investigates the accuracy of this single-level approximation. The single-level approximation, of course, cannot be used to explain how the learning and forgetting occur at various levels of an organization. However, numerical experiments based upon the Liberty ships dataset show that the single-level approximation can provide surprisingly good estimates of the organization's key performance measure, e.g., production time per unit. It can therefore yield good estimates of the learning and forgetting rates aggregated for the entire organization, and these estimates can be used to compare the performance of one organization to another. The single-level approximation is shown to perform particularly well when the data exhibit a large amount of dispersion, the number of units used for fitting is large, the learning occurs slowly, or the forgetting rate is high.

1. Introduction

Organizational learning is a well-known phenomenon evidenced by numerous empirical studies. Since Wright's (1936) application of the log-linear learning curve function to aircraft manufacturing, many different forms of learning curve functions have been proposed and examined in various types of processes across different industries. Yelle (1979) provided an extensive survey of the classical learning curve functions published in the management and engineering literature. Jaber (2011) provided a collection of recent developments in learning curve models and their applications in the area of management, economics, engineering, and psychology. Nembhard and Uzumeri (2000) empirically compared various functional forms of the learning curve. Balasubramanian and Lieberman (2011) estimated learning curve rates in over 250 U.S. industries, finding a wide range of rates between and within industries. More recently, Grosse, Glock, Christoph, and Müller (2015) fit eleven different learning curve functions to over a hundred datasets to determine which functions obtained the best fit for different types of data. For many industries, learning models have been incorporated into broader planning models in pursuit of better resource allocation decisions (Grosse, Glock, & Christoph, 2015a,2015b; Nadeau, Kar, Roth, & Kirchain, 2010; Nembhard & Bentefouet, 2015; Van Peteghem & Van Houcke, 2015). Given the large variations in learning rates across and within industries, a key element of successful planning is determining an accurate estimate of future learning.

The reverse of learning, i.e., forgetting, has also been recognized at the organizational level and incorporated into learning curve functions. Argote, Beckman, and Epple (1990) developed a discrete time forgetting model and applied it to the construction of Liberty ships during World War II. Other authors applied similar models to other industries (Benkard 2000; Darr, Argote, & Epple, 1995; Epple, Argote, & Murphy, 1996). Using a new, disaggregated dataset for the Liberty ship data, Thompson (2007) developed a forgetting model in which knowledge was assumed to depreciate continuously over time. While Thompson's model used a single time series for each Liberty shipyard, Kim, Moskowitz, Plante, Seo, and Tang (2007) and Kim and Seo (2009)

* Corresponding author.
E-mail addresses: ilhyung.kim@wwu.edu (I. Kim), mark.springer@wwu.edu (M. Springer), george.zhang@wwu.edu (Z.G. Zhang), yspark@skuniv.ac.kr (Y.-S. Park).

https://doi.org/10.1016/j.cie.2018.10.004

Please cite this article as: Kim, I., Computers & Industrial Engineering, https://doi.org/10.1016/j.cie.2018.10.004

extended Thompson's model by considering multiple time series for each shipyard, one for each of the multiple shipways in a shipyard. Just as learning models have been used in a broad range of resource planning and allocation decisions, the notion of forgetting has also been incorporated into various planning models (Heimerl & Kolisch, 2010; Suer & Tummaluri, 2008).

While the above authors have focused on measuring the amount of learning and forgetting, other authors have speculated on the possible components of organizational learning. Recently, Glock and Jaber (2014) proposed a mathematical model in which group or organizational learning develops based on the compatibility of group members' knowledge, their willingness and ability to share and absorb knowledge, and the structure of the group. Without speculating on the precise relationship of individual and organizational learning, Harvey (1979) suggested a breakdown of "total [organizational] learning" into two constituents: learning at the operator level was due to increased worker dexterity and task competency at each work station, while higher-level learning consisted of improved task definition and process management. Similarly, Argote (1999) grouped organizational learning factors into three categories: increased individual proficiency, improvements in the organization's technology, and improvements in its structure, routines and methods of coordination. This suggests that learning and, by extension, forgetting occurs at every organizational level and that the rates of learning and forgetting vary from one level to another.

Most empirical studies, however, measure learning and forgetting at a single aggregated level of an organization. These studies therefore implicitly assume that the estimation obtained from this single-level approach is a good approximation for the sum of all learning and forgetting occurring at the various levels of the organization. If different learning and forgetting rates were in effect at different levels in an organization, however, this approximation could lead to poor performance measure estimates and, consequently, poor managerial decisions.

Recently, Park, Springer, and Kim (2013) proposed a model which considers both learning and forgetting at two levels in an organization. They applied the model to the Liberty ships dataset, since each Liberty shipyard possessed a two-level organization: each "shipyard," or facility, consisted of one or more "shipways," or production lines. Learning and forgetting was measured at both the shipway and the shipyard levels. They found that almost all of the shipway-level learning occurred during the production of the first few ships, while the shipyard-level learning occurred throughout the production of all ships. Based on these results, they theorized that the knowledge at the shipway level was stored in a soft form manifested in the operational skills of workers and, as a result, was acquired and depreciated at a faster rate, while the knowledge at the shipyard level was stored in a hard form such as documents and, therefore, was acquired and depreciated at a slower rate.

When learning and forgetting occur differently at the various levels of an organization, a multiple-level model has the potential for greater accuracy than a single-level model in estimating the amount of learning and forgetting in the organization. However, use of a multiple-level model is not without drawbacks. The multiple-level model has at least twice as many parameters as the single-level model; in addition, the multiple-level model requires a disaggregated dataset, which allows the researcher to track each unit produced through the multiple levels of the organization. The multiple-level model may therefore not be a feasible option for every dataset because of the depth of detail that the model requires. Consequently, it is important to determine how much the two models differ in their prediction accuracy and under what circumstances.

In this paper, the authors compare the performance of the single- and two-level models. Specifically, we first generate datasets based on a two-level learning model, i.e., we assume an organization exhibits both learning and forgetting at two levels, an upper and a lower level, of the organization. Analogous to the Liberty shipbuilding process, the upper level could represent a factory and the lower level could represent each of the multiple production lines in the factory. The range of learning and forgetting exhibited in the datasets is anchored in the parameter estimates obtained from the earlier analysis of the Liberty ship datasets (Park et al., 2013).

We then fit the generated datasets using two alternative single-level models as well as the two-level model; one single-level model assumes that all learning takes place at the factory-wide organizational level, while the other single-level model explicitly considers only the learning that happens within each production line. The experiment therefore mimics situations where multiple levels exist in an organization, but due to a paucity of data, the difficulty of data collection, or a lack of modelling sophistication, a single-level learning and forgetting model is applied. While the two-level model can be expected to outperform both single-level approximations under such circumstances, the relative performance of these single-level approximations and the factors affecting their performance are the subjects of this study. Given the common use of single-level models in studies of learning and forgetting, answers to these questions have important managerial implications.

The rest of the paper is organized as follows: In Section 2, the details of the multiple-level learning and forgetting models are presented. In Section 3, the Liberty ships dataset and the methodology for generating datasets is discussed. In Section 4, we compare the performance of the single- and the two-level models and discuss the accuracy of the single-level approximations. Section 5 provides conclusions.

2. Multiple-level learning and forgetting models

2.1. A modeling framework of multiple-level learning and forgetting models

Consider an organization with M organizational levels. At each level, there exists a multiple number of parallel entities. As an example, consider an automobile manufacturer which produces a vehicle at multiple plants, each of which consists of multiple and parallel assembly lines. This manufacturer is an example of a three-level model ($M = 3$) with the company, plants, and assembly lines being the first, second, and third levels. The total effect of learning and forgetting is the product of the learning and forgetting at each level, a functional form consistent with the Cobb-Douglas production function (Cobb & Douglas, 1928).

The mathematical representation of a multiple-level learning and forgetting model is then as follows:

$$Q_{[k_M]} = \alpha \prod_{m=1}^{M} F^m(E^m_{[k_m]}) e^{\varepsilon_{[k_M]}}, \;\; \varepsilon_{[k_M]} \sim Normal\ (0, \sigma), \tag{1}$$

where $Q_{[k_M]}$ is the performance measure, *e.g.*, the processing time or production cost, for the k^{th} item produced in level M. In order to apply a multiple-level model, the dataset containing all items produced in the organization needs to be arranged M different ways, one for each level. At the first or the highest level, all items should be configured into a single time series ordered by the start or completion time of each unit. At the second level, the same items should be configured into multiple parallel time series, one for each organizational entity at the second level, ordered by the start or completion time of each unit in the corresponding entity. For notational simplicity, a single index in a bracket $[k_m]$ is used to represent the ordering for level m. In the above equation, the performance measure $Q_{[k_M]}$ is specified by the index at level M, the lowest organizational level in an M-level model. However, it could be indexed by any time series at any level since the items in the dataset are identical no matter how they are indexed.

The constant α represents the initial starting point of the performance measure. The learning curve function $F^m(\cdot)$ at level m determines how the performance measure is affected by the stock of experience or knowledge $E^m_{[k_m]}$ available for the $[k]$th unit at level m. Finally, $\varepsilon_{[k_M]}$ is an error term for the $[k]$th unit at level M, which is an identically and independently distributed normal random variable with a mean of 0 and standard deviation of α.

As the production continues, the stock of knowledge accumulates. On the other hand, as time passes, this stock continuously depreciates. To capture these phenomena, we assume that the stock of experience or knowledge at each level accumulates and depreciates according to the following recursive mechanism:

$$E^m_{[k_m]} = e^{-\delta_m S_{[k_m]}}(E^m_{[k_m-1]} + 1), \text{ for } m = 1, 2, \ldots, M \tag{2}$$

The interpretation of the above recursive formula is straightforward. First, the accumulation of knowledge is shown within the parenthesis. The unit of knowledge is defined such that the production of an item adds one unit of knowledge to its stock at each level. It should be noted that this does not mean that the same amount of knowledge is accumulated at all levels. By assuming independent and distinctive parameters of learning and forgetting for each level, the model allows the production of an item to generate different amounts of knowledge for each level. Second, the stock of knowledge depreciates from one item to the next due to the time elapsed between the two items. Specifically, $S_{[k_m]}$, the difference in production starting time between the $[k]$th and $[k-1]$th items at level m is used in Eq. (2). The instantaneous knowledge depreciation rate at level m is represented by δ_m. Thus, as the production continues, the stocks of knowledge may increase or decrease depending on whether the acquisition of new knowledge is greater or less than the depreciation of knowledge.

It is assumed that the forgetting occurs continuously during the elapsed time between two consecutive units. At the individual level, it would be more plausible to assume that forgetting begins immediately upon ceasing production but not during production (Jaber, 2006). At the organizational or group level, on the other hand, it is often assumed that the forgetting occurs throughout the period, including a break as well as a production period (e.g., Sikstrom & Jaber, 2012). Regarding a functional form for forgetting, an exponential function is used in Eq. (2). As Bailey (1989) discovered in his experiments, forgetting mainly depends on the passage of time and the stock of knowledge. We therefore adopt a continuous depreciation of knowledge. Specifically, the stock of knowledge E depreciates over time t at a constant rate δ depending on the stock of knowledge: $dE/dt = -\delta E$. By solving the above differential equation, we have $E(t) = E_0 e^{-\delta t}$, an exponential decay, which is common in many physical phenomena (e.g., radioactive decay) and also in the organizational forgetting examined in many empirical studies (Argote et al., 1990; Benkard, 2000; Kim & Seo, 2009; Kim et al., 2007; Park et al., 2013; Thompson, 2007). While our current model is restricted to exponential forgetting, it should be noted that the power function, as a mirror image of a power learning function, has also been commonly used to model forgetting (Sikstrom & Jaber, 2012).

The multiple-level model given in Eq. (1) does not contain an explicit functional form of the knowledge transfer (or interactions) between different organizational levels or among the entities in the same organizational level. However, the model considers such interactions through the stocks of knowledge stored at multiple levels. Consider a two-level case, where the upper level organizational entity is a plant and the lower level organizational entities are assembly lines. As shown in Eq. (1), the performance measure of a unit produced in an assembly line depends not only on the stock of knowledge accumulated at the assembly line, but also on that accumulated at the plant. At the same time, the stock of knowledge at the plant depends on any units produced in any assembly lines as well as its own creation of knowledge at the plant as shown in Eq. (2). In other words, the knowledge created at an assembly line will be transferred to the plant and then transferred back down to another assembly line through the mechanisms specified in Eqs. (1) and (2). These equations can therefore be used to model very different relationships between organizational levels. For example, the plant may invest in a formal "skunkworks" that repeatedly designs and rolls out process improvements to all assembly lines; if the assembly line work was highly standardized and designed to require limited worker judgement, learning at the upper level by management would be much more than the sum of the learning occurring on the assembly lines spread across the organization. Alternatively, the plant could have minimal formal upper-level process improvement initiatives, and instead rely on accumulating improvements developed by experienced workers on the shop floor. In such a situation, upper-level learning would largely be a delayed reaction to the learning at the lower levels of the organization. And it is, of course, possible to imagine situations between these two extremes.

The multiple-level learning and forgetting model presented here does not assume a functional relationship between learning or forgetting at the different levels, but rather relies on the timing of observed learning and forgetting between as well as within lower organizational units to estimate the contribution to learning and forgetting at each level. The nature of the relationship between learning and forgetting at the different levels may be of interest, but our model only estimates the rates (and amounts) of learning and forgetting at each level, not the extent of any functional relationship. The model therefore has the advantage of being applicable regardless of the relationship between learning and forgetting at the different organizational levels.

2.2. Learning curve functions

Three learning curve functions are considered for our experiment: log-linear, accumulation, and replacement functions. While the log-linear function implicitly assumes unbounded learning, the accumulation and the replacement functions are bounded, i.e., the performance measure cannot be improved beyond a certain terminal value. Conceptually, the accumulation function assumes that new knowledge is added without loss of prior knowledge; the replacement function assumes that some knowledge is replaced by other more productive knowledge (Restle & Greeno, 1970). The functional forms of the learning curve for all levels are assumed to be identical. The log-linear function has the forms

$$F^m(E^m_{[k_m]}) = (E^m_{[k_m]} + \kappa_m)^{-\beta_m}, \text{ for } m = 1 \text{ or } 2. \tag{3}$$

The accumulation function has the forms

$$F^m(E^m_{[k_m]}) = \frac{1 + \tau_m \beta_m E^m_{[k_m]}}{1 + \beta_m E^m_{[k_m]}}, \text{ for } m = 1 \text{ or } 2. \tag{4}$$

Finally, the replacement function has the forms

$$F^m(E^m_{[k_m]}) = \tau_m + (1-\tau_m)(1-\beta_m)^{E^m_{[k_m]}}, \text{ for } m = 1 \text{ or } 2. \tag{5}$$

β_m is the rate of learning at level m. Consider a case when the performance measure is either production time or cost. As the stock of knowledge approaches infinity, the log-linear curve asymptotically converges to zero, while the accumulation and the replacement curves converge to the terminal value τ_m at level m. For the log-linear function, κ_m is the initial stock of knowledge at level m. This initial stock is needed in the log-linear learning function to avoid the initializations problem of the first unit in each level. In the numerical experiments discussed in section 3, we set all initial knowledge stocks equal to one without loss of generality.

2.3. Parameter estimation

To estimate the parameters of the above model, the parameters of M different time series must be estimated simultaneously, where all time series are constructed from a single dataset but arranged in M different ways. For the experiments in this paper, a two-level organization was modelled, so an iterative algorithm which linked the values in the two time series was used. Details of the algorithm may be found in Park et al. (2013).

2.4. The single-level approximations

We consider two different versions of the single-level model. The first single-level model assumes that learning and forgetting occur only at the upper level of the organization, while the second single-level model assumes that they occur only within the lower level organizational entities. Most of the existing research in learning and forgetting relies on aggregation at the upper-level (Argote et al., 1990; Benkard, 2000; Darr et al., 1995; Epple et al., 1996; Thompson, 2007), while only a handful of researchers utilize aggregation at the lower-level (Kim & Seo, 2009; Kim et al., 2007). We refer to the former as the upper single-level model and the latter as the lower single-level model. The upper single-level model considers a single and long sequence of units produced in the entire organization, while the lower single-level model considers multiple short sequences of units produced in each of the lower level entities separately. The long sequence of the upper single-level model contains the order information of the units in each of the short sequences of the lower single-level model. As a result, it is likely that the upper single-level model will perform better than the lower single-level model unless most of the learning occurs at the lower level and little or no learning occurs at the upper level.

It is important to emphasize that we are considering the relative accuracy of single-level models when used as approximations in a situation where the learning and forgetting are actually occurring at multiple levels. Large organizations are likely to have multiple organizational levels, but, as noted above, most learning and forgetting research focuses on the application of single-level models. Given the frequency with which single-level models are used to approximate learning and forgetting at multiple organizational levels, and the existence of models (e.g., Park et al., 2013) that can estimate learning and forgetting at multiple levels, it is important to know what – if anything – is being lost by continuing to use single-level approximations when learning and forgetting are taking place at multiple levels. More precisely, we wish to answer three key questions. First, are multiple-level models more accurate in forecasting production time than single-level models when learning and forgetting occur at multiple levels? The answer to this question is not necessarily yes, since the number of parameters in multiple-level models can be quite large compared to single-level models; this makes it difficult for the multiple-level models to provide accurate forcasts, especially when the system exhibits a large amount of variability. Second, if the answer to the first question is yes, how large is the difference in accuracy? And finally, how do different factors – such as production time variability, or the underlying learning and forgetting rates at the different organizational levels – impact the relative performance of the multiple-level and single-level models? Even if a difference exists "on average" between models, it is likely that the magnitude of the difference varies depending upon the nature of the data. Knowing how external factors impact model performance can help guide model selection in a real-world application.

3. Data generation for numerical experiments

3.1. The liberty ships case

To generate data close to a real situation, the benchmark dataset of Liberty ships is used as a basis for determining the parameter ranges. During World War II, a total of 2710 Liberty ships were produced at sixteen shipyards. Most shipyards[1] produced a large number of ships, ranging from 82 to 384 ships for a period of two to four years. Each shipyard consisted of multiple parallel shipways ranging from 6 to 16 shipways, and each shipway produced between 13 and 31 Liberty ships. All shipyards experienced a great amount of learning, with production time per ship dropping from as high as 332 days to as low as 13 days.[2] This dataset has a unique structure which is not commonly found in other datasets. A single homogeneous product, the Liberty ship, was produced from homogenous raw materials in a large number of organizations (sixteen different shipyards) with workers who shared a common level of prior industry experience. Consequently, the Liberty ships case has been widely used in many previous studies on learning as mentioned in Section 1.

Park et al. (2013) measured learning and forgetting at both the shipway and the shipyard level by applying their two-level model, finding that learning at the shipway level occurred at a much faster rate than learning at the shipyard level. We use the estimation results of the Liberty ships case, discussed below, as a foundation in determining the parameter values of our two-level model.

3.2. Shipyard and shipway size

In the Liberty ships case, the average number of shipways per shipyard was 9.69, and the largest number of ships produced in a single shipway was 31. The organization modelled in this paper therefore has ten parallel lower-level organizational entities, each of which can produce at most 30 units of product, resulting in a total of 300 units for the entire organization.

3.3. Learning rates and the amount of learning

For the log-linear function, a single parameter β_m at each organizational level determines both the rate and the amount of learning. For the accumulation and the replacement functions, however, an additional shape parameter τ_m, the terminal value, allows the learning amount to be manipulated independently of the learning rate.

3.3.1. Amount of learning: high vs. low

To enable a fair comparison of all three learning curve functions, it is first necessary to determine what constitutes a "high" and a "low" amount of learning at each organizational level; parameters for the three functions can then be chosen such that the amount of learning after a specified number of units is approximately the same for each function at a specified organizational level of learning. The "amount of learning," υ_m, is defined as a fraction indicating the proportionate reduction in the production time of a unit of product after the production of a fixed number of units. In the Liberty ships data, the observed log-linear learning rates implied learning amounts of 46.5% at the upper level and 28.7% at the lower level after production of 300 units. These learning amounts were similar to those obtained from the two bounded learning functions, which were estimated as 0.529 and 0.264 for the replacement function and 0.444 and 0.196 for the accumulation function. Since the accumulation function requires a very large number of units until it reaches its terminal values, while the replacement function reaches its terminal values rather quickly, it was considered more straightforward to first set the parameters of the replacement function and then determine those of the accumulation function to make the accumulation function learning curve achieve a certain target amount of learning within a given number of units. This paper therefore, based on the replacement function, sets 0.3 and 0.5 as the values of high and low amounts of learning achieved by the time of the production of the last unit.[3] These values are shown as the upper- and lower-level

[1] Three of the sixteen shipyards produced less than 20 Liberty ships and were converted to produce other types of ships after a short period of producing Liberty ships.

[2] The production time of one ship produced at the second Permanente Metals-Richmond shipyard in Richmond, CA, was only 7 days, which in fact was set up for propaganda and was not considered as the shortest one.

[3] Since the learning occurs at two levels, the total amount of learning combining the two levels together will be much higher. As an example, if one level has high learning (terminal value = 0.3) and the other level has low learning

Table 1
Parameter values for experiment.

Learning curve function	Experimental factors	Factor notation	Corresponding parameter (If different from factor)	Parameter notation (If not factor)	Low value	High value
All functions	Initial # of units per production line	n			50	150
	Standard deviation of error	σ			0.010	0.050
	Upper-level forgetting rate	δ_1			0.100	0.200
	Lower-level forgetting rate	δ_2			0.100	0.200
Log-linear	Upper-level learning rate	β_1			0.121	0.211
	Lower-level learning rate	β_2			0.204	0.354
Accumulation	Upper-level learning rate	β_1			0.009	0.043
	Lower-level learning rate	β_2			0.093	0.533
	Upper-level learning amount	ν_1	Upper-level terminal value	τ_1 (with low β_1)	0.326	0.056
				τ_1 (with high β_1)	0.439	0.215
	Lower-level learning amount	ν_2	Lower-level terminal value	τ_2 (with low β_2)	0.330	0.062
				τ_2 (with high β_2)	0.448	0.227
Replacement	Upper-level learning rate	β_1			0.010	0.030
	Lower-level forgetting rate	β_2			0.100	0.300
	Upper-level learning amount	ν_1	Upper-level terminal value	τ_1	0.500	0.300
	Lower-level learning amount	ν_2	Lower-level terminal value	τ_2	0.500	0.300

terminal values, τ_1 and τ_2, of the replacement function in Table 1.

3.3.2. Rate of learning: fast vs. slow

The amount of learning υ_m determines how much learning will occur, while the learning rate β_m determines how fast the maximum amount of learning can be achieved. As mentioned above, the learning amount and learning rate are not independent for the log-linear function. Thus, if the learning function is to yield learning amounts of 0.3 for high learning and 0.5 for low learning, this will determine the values of the learning rate parameters for fast and slow learning. The values of the upper and lower learning rates β_m that yield these learning amounts for the log-linear function are shown in Table 1. When these rates are translated into "progress ratios," i.e., the percentages that the performance measure is reduced to by a doubling of cumulative output, the range of rates is quite similar to the range found by Balasubramanian and Lieberman (2011) in their aforementioned survey of 250 industries; the twenty-fifth and seventy-fifth percentiles of progress ratios across all examined facilities were found to be 87% and 78%, respectively.

For the replacement and accumulation functions, since the learning rates are independent of the learning amount, the impact of fast vs. slow learning rates can also be investigated. Park et al. (2013) observed that almost all of the shipway-level learning occurred during the first few ships while the shipyard-level learning occurred throughout the production of all ships. The fast and slow learning rates can be set to show the similar patterns. Specifically, with the fast learning rate, approximately 95% of learning can be achieved by the first one third of the units (the 100th and 10th units for the upper and lower levels, respectively), while, with the slow learning rate, the same amount of learning can be achieved by the last unit (the 300th and 30th units for the upper and lower levels, respectively). Learning rates at the lower and upper levels of the replacement function are therefore set at 0.3 and 0.03, respectively, for fast learning, and 0.1 and 0.01, respectively, for slow learning. Since the organization (upper level) has ten parallel sub-organizational entities (lower levels), whenever each sub-entity produces a unit, the entire organization produces, on average, ten units. That is why the learning rate of the upper level is only one tenth of that of the lower level for the same speed of learning.

Finally, the parameters of the accumulation function are determined so as to make its learning curves as close as possible to those of the replacement function for a given learning rate and learning amount combination. A simple optimization approach is used to minimize the difference between the learning curves of the two functions. Note that for the accumulation function, the terminal value τ_m needed to achieve a certain amount of learning depends on the value of the learning rate as well as the desired amount of learning. Hence, as shown in Table 1, the parameters τ_1 and τ_2 necessary to calculate the learning function are determined by, but are distinct from, the experimental factors υ_1 and υ_2 used to investigate the impact of the learning amount. For the replacement function, on the other hand, τ_1 and τ_2 are equivalent to υ_1 and υ_2.

3.4. Factors common across all learning functions

In addition to the two learning curve function-specific factors discussed in the previous sections, we consider another important factor, the forgetting rate, which clearly affects the total amount of learning over the course of production. The Liberty ships case demonstrated that learning was not persistent but depreciated over time; the forgetting rate was high at the shipway level and low at the shipyard level. Based on the these results, the instant forgetting rate, δ_m, was set at 0.2 for the high level of forgetting and 0.1 for the low level of forgetting; these correspond to monthly knowledge retention rates of 90.4% and 81.9%, respectively. To further assess the robustness of the single-level approximations, the level of randomness and the number of production units available for model fitting are also introduced as factors. The higher the degree of randomness in the data, the more difficult it is likely to be to estimate the parameters. As a result, the forecasting accuracy is expected to deteriorate as the data contain more noise.

The standard deviation of the error term, σ, is set at two levels: 0.01 for a low level of dispersion and 0.05 for a high level of dispersion. In Park et al. (2013), the estimated values of σ are 0.105, 0.119, and 0.116 for the two-level log-linear, accumulation, and replacement models fitted to the liberty ship data. The levels of sigma used in the current project are therefore lower than those estimated for the liberty ship data. As discussed in greater detail below in Section 4.2, the lower range for σ increased our ability to identify those factors which differently impacted the performance of our learning and forgetting models.

The final factor is the number of units, n, which is available for fitting the learning curve functions. As a greater number of units are observed, the parameter estimates of the learning curve functions should become more accurate. Two cases are tested: when data for the

(*footnote continued*)
(terminal value = 0.5), the terminal value of the total learning will be 0.3 × 0.5 = 0.15, i.e., the production time converges to 15% of the initial production time as the cumulative number of units approaches to infinity.

Table 2
Mean Absolute Percentage Errors for different learning curve functions.

Learning curve function	Forecasting model	Number of observations	Mean MAPE	Median test: Kruskal-Wallis significance	Paired sign test significance		
					Forecasting model		
					Two	Upper	Lower
Log-Linear	Two	192	0.0355	8.41E-20		2.11E-11	1.18E-28
	Upper	192	0.0452				6.86E-15
	Lower	192	0.0691				
Accumulation	Two	768	0.0361	7.88E-64		2.49E-32	1.21E-87
	Upper	768	0.0435				1.24E-41
	Lower	768	0.0715				
Replacement	Two	768	0.0487	6.75E-49		1.80E-18	5.60E-72
	Upper	768	0.0641				5.87E-60
	Lower	768	0.1210				

first 50 units are available, and when data for the first 150 units are available. In both cases, these are the first units available from the perspective of the upper organizational level, which corresponds roughly to the first 5 units and the first 15 units per each organizational sub-entity at the lower level. Parameters for the learning curve functions are estimated based on the above initial observations, and then production times are forecasted for the next 100 units. Accuracy of the forecasts is measured using the mean absolute percentage error (MAPE).

3.5. Generation of data sets

A complete list of factors, as well as the corresponding parameter values necessary for each factor combination, is given in Table 1. Datasets are generated based on a two-level full factorial design with three replications (Box, Huner, & Hunter, 1978); note that in this context, “two-level” refers to the number of levels for each factor in the experiment, and does not refer to the number of levels of learning in the organization. For all generated datasets in the experiments, for example, the forgetting rate at the upper organizational level was set to either 0.10 or 0.20 Relying on a two-level full factorial design ensures that the estimates of the regression coefficients are statistically independent, i.e., there is no multicollinearity. Thus, the estimate of the impact of the forgetting rate at the upper organizational level will not be “confounded” with the estimate of the impact of any other factor in the experiment. A single set of factors, with each factor set at one of the two factor levels specified in Table 1, is therefore the basis for each dataset. For a single replication of a two-level full factorial design, each dataset’s factor combination is unique. Using the standard procedure for generating two-level full factorial designs (Box et al., 1978), with six factors the total number of unique factor combinations generated for a single replication of the log-linear function experiment is 2^6 or 64. For the replacement and accumulation function experiments, we have two additional factors, and the total number of unique factor combinations generated for a single replication of each of these experiments is therefore 2^8 or 256.

In generating each dataset, once the parameter settings are fixed the only random element is the error term $\varepsilon_{[k_M]}$ specified in (1), and the size of the error terms is of course dependent on the magnitude of the factor σ. If σ is nonzero, as it is in all of our experiments, the same unique factor combination with a different random number seed will generate different datasets. To improve our estimates of MAPE, as well as to provide a measure of standard error to enable hypothesis testing, each experiment consisted of three replications, i.e., each unique factor combination was used to generate three different datasets using three different random number seeds. This resulted in 64 × 3 or 192 datasets for the log-linear function experiment and in 256 × 3 or 768 datasets for the accumulation and replacement function experiments. For the regressional analysis discussed below in Section 4.2, the factor settings (low or high) for each dataset and the MAPE of the resulting model fit constituted a single observation, and the regressions were therefore performed on either 192 (log-linear) or 768 (accumulation and replacement) observations.

4. Forecasting accuracy of single-level approximation

Clearly, using a single-level learning model for an organization with multi-level learning is not appropriate if our primary interest is in the accurate estimation of learning and forgetting rates at each level. However, if the primary interest is in predicting the performance measure, which is usually a productivity measure such as production costs, manufacturing lead times or labor hours, the aggregated approach based on a single learning curve may provide reasonably accurate estimates

4.1. Overall performance of single-level approximations

For each dataset generated as described in Section 3.5, three different versions of the corresponding learning function were fitted: the two-level model, the upper single-level model, and the lower single-level model. We refer to these three versions of curve-fitting models as *forecasting* models. For each fitted dataset, the MAPE for the next 100 production estimates was then calculated. Table 2 shows the mean MAPEs for all nine combinations constructed by three learning curve functions (log-linear, accumulation, and replacement) and three forecasting models (two-level, upper-level, and lower-level).

As one can see from the table, the pattern for each of the three different learning curve functions is the same: the two-level model outperforms the upper-level model, and the lower-level model has the worst performance. It is certainly not surprising that, on average, the two-level models outperform the two single-level models, as two-level learning was the basis of each dataset; any errors observable for the two-level model therefore represent error introduced by the randomness of the data.[4] Perhaps surprisingly, however, the mean MAPE for the upper-level models is generally greater than that for the two-level models by only 1 or 2%, while the lower-level models generally have MAPEs approximately twice as great (greater by 3 to 8%) as the two-level MAPEs.

To formally test for significant differences among the MAPEs of the three forecasting models, a Kruskal-Wallis medians test[5] was performed

[4] The two-level models consistently yielded a MAPE of zero when the underlying data were deterministic.

[5] The distribution of MAPEs was highly skewed to the right for all nine combinations (three learning curve functions × three forecasting models) and far from a normal distribution. Thus, we used this non-parametric test instead

Table 3
Impact of experimental factors on MAPE.

		Effect estimates								
		Log-linear function			Accumulation function			Replacement function		
Factor	Factor notation	Two-level	Upper level	Lower level	Two-level	Upper level	Lower level	Two-level	Upper level	Lower level
Initial # of units per production line	n	*−0.0078*	*−0.0111*	*−0.0234*	*−0.0087*	*−0.0128*	*−0.0244*	*−0.0192*	*−0.0288*	*−0.0526*
Standard deviation of error term	σ	*0.0198*	*0.0117*	*0.0097*	*0.0203*	*0.0151*	*0.0076*	*0.0269*	*0.0159*	*−0.0070*
Upper-level learning rate	β_1	*0.0035*	*0.0067*	*0.0220*	*0.0034*	*0.0088*	*0.0100*	*0.0099*	*0.0219*	*0.0350*
Lower-level learning rate	β_2	*0.0019*	*0.0025*	*0.0173*	*0.0009*	*0.0016*	*0.0222*	*0.0053*	*0.0143*	*0.0615*
Upper-level learning amount	ν_1				*0.0024*	*0.0034*	*0.0142*	*0.0067*	*0.0163*	*0.0312*
Lower-level learning amount	ν_2				*0.0010*	0.0000	*0.0179*	*0.0045*	*0.0101*	*0.0382*
Upper-level forgetting rate	δ_1	*−0.0022*	*−0.0026*	*−0.0097*	*−0.0027*	*−0.0043*	*−0.0074*	*−0.0063*	*−0.0163*	*−0.0176*
Lower-level forgetting rate	δ_2	0.0010	*0.0030*	*−0.0063*	−0.0008	−0.0010	*−0.0113*	*−0.0052*	*−0.0055*	*−0.0262*
Adjusted R-squared		0.8711	0.6118	0.9117	0.7815	0.6631	0.8975	0.5431	0.6722	0.7484

Italicized numbers are significant at the 0.05 level.

for each learning curve function. As can be seen in Table 2, the null hypothesis of equality among the median MAPEs of the forecasting models was rejected for each of the three learning curve functions with extremely low p-values. For each learning curve function, this result was followed up with three paired sign tests for the equivalence between the median MAPEs for each pair of the forecasting models. Table 2 again shows that the null hypotheses of equivalence are rejected in each instance with a high degree of significance.

While the differences between the different forecasting models are statistically significant, however, they are not necessarily of practical and managerial significance. The difference in mean MAPE between the two-level and the upper-level model is 0.97% for the log-linear function, 0.74% for the accumulation function, and 1.54% for the replacement function. In practical terms, the upper-level model may be considered to perform equally well as the two-level model in terms of average forecasting accuracy.

4.2. Impact of learning parameters on model accuracy

To examine how MAPEs behave with respect to the experimental factors, for each of the nine learning curve function/forecasting model combinations the MAPE was regressed on the factors. The results are given in Table 3. The values reported for each factor are the increase (+) or decrease (−) in the MAPE resulting from setting the factor at its high level; the additive inverse of the reported value would reflect the impact on the MAPE of setting the factor at its low level. For example, setting the initial number of units n used to fit the data at its high value of 150 *reduces* the MAPE, on average, by 0.78% from the overall mean MAPE of 3.55% for the two-level model using log-linear learning functions; setting n to its low value of 50 would *increase* the MAPE by 0.78% over the average mean MAPE for this model. Italicized estimates are significant at the 0.05 level.

Not surprisingly, the coefficients of the factor for n are negative for all nine learning function/forecasting model combinations, i.e., as n increases, the MAPE decreases. Perhaps less obviously, the two-level models appear to be least affected by the amount of data available for fitting, while the availability of data more dramatically affects the lower-level models. This effect is evident in Fig. 1, which illustrates the relationship between the MAPE and n for a greater range of n. Fig. 1 shows the MAPEs when fitting two-level, upper-level, and lower-level models to data from a two-level learning system with the log-linear learning curve function. The parameters for the underlying two-level learning systems have been chosen to make the data difficult to fit. Each point plotted on the graph represents an average of thirty observations.

While all models show improvement as n increases from five to twenty, the more poorly-performing single-level models shows a greater sensitivity to data availability. Since, for a specified rate and total amount of learning, the per-unit amount of learning is greatest for items early in the learning process, this phenomenon likely indicates that the more accurate two-level model loses its comparative advantage as the learning curves at both levels flatten out.

Another expected phenomenon is the change in MAPE with respect to the standard deviation σ of the disturbance term; as σ increases, the MAPE also increases. In contrast with the factor n, σ appears to have a greater impact on the accuracy of the two-level models than the single-level models. As shown in Fig. 2, increasing σ results in an increase in MAPE for all three models, but the performance of the single-level models deteriorates less rapidly than that of the two-level model. The declining advantage of the two-level model when σ is large may be due in part to the number of parameters in the two-level model; since the two-level model has twice as many parameters as either single-level model, as σ increases the standard errors associated with the estimators also increase faster. This phenomenon was investigated further in a series of follow-up experiments where σ was set at 0.10 for all datasets. For each model and learning function combination shown in Table 4, an experimental run was conducted using the same factor levels and number of replications as discussed above, except that sigma was set at 0.10 for all trials. The resulting learning model MAPEs and adjusted R-squared values of the regression models examining the impact of each factor are shown in columns 3 and 4 of Table 4; the MAPEs and adjusted R-squared values for the initial experiments shown in Table 2 are reproduced in columns 5 and 6 for comparison.

As expected, the MAPEs for the high-sigma experiments are much higher than the MAPEs for the original (sigma ranging from 0.01 to 0.05) experiments. Equally important, however, is that at the higher level of sigma, the difference between the upper-level and two-level model is negligible or at least much reduced. This is the logical extension of the phenomenon noted in Fig. 2. Since our thesis is that the upper-level model can, under certain circumstances, perform as well as the two-level model even when the actual learning and forgetting are taking place at two levels, and we did not want to overly bias the experiments in favor of our thesis, we chose parameters settings that gave the two-level models the best chance of outperforming the single-level models.

In addition, a sigma value of 0.10 introduces variability that limits the ability of the regression model to predict the impact of each factor; not only are the MAPEs higher for the high sigma experiments, but the adjusted R-squared values of the regressions are also notably lower. Since one goal of the current project was to judge the impact of the factors on model accuracy, with an emphasis on identifying factors affecting the relative performance of two-level and single-level models, the authors were concerned that using a higher sigma (e.g., 0.10) for

(*footnote continued*)
of single-factor ANOVA.

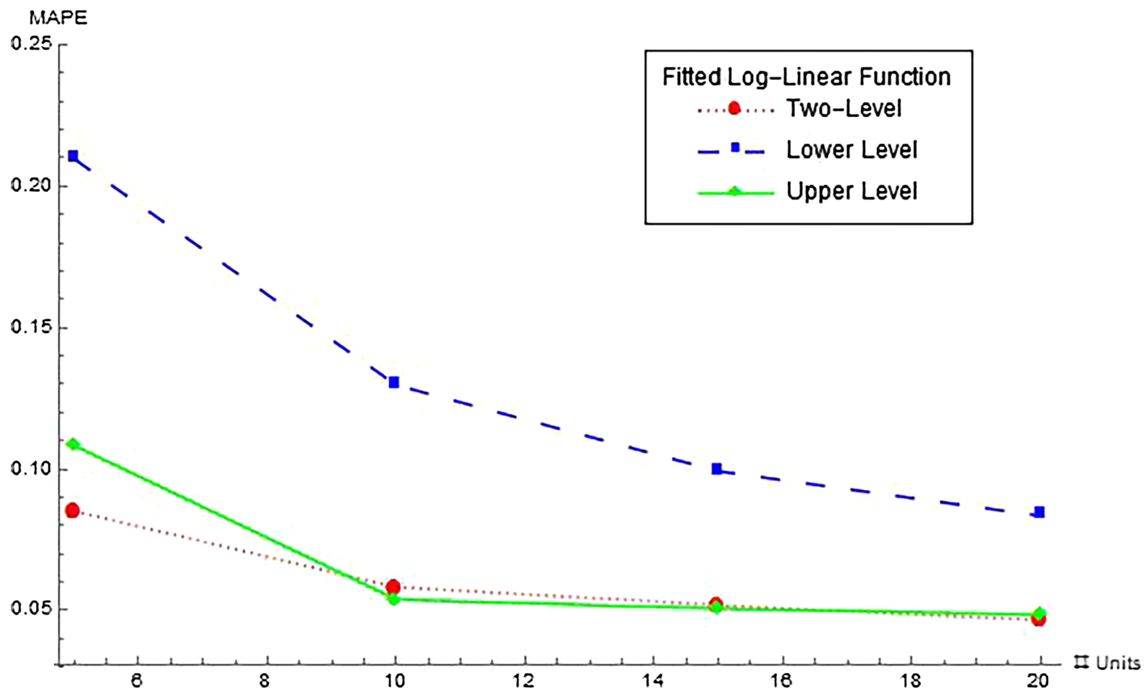

Fig. 1. MAPE for log-linear function for various levels of n.

the high factor setting would introduce so much variability that it would mask the impact of the other factors.

All six remaining explanatory variables also show a consistent behavior. The upper and lower-level learning rate factor effects are significant and positive in all nine learning function/forecasting model combinations, and the impact on MAPE is greatest for the lower-level models and smallest for the two-level models; a high learning rate leads to poorer forecasts, and this is more so the case for the single-level approximations. Eleven of the twelve learning amount factor effects are significant and positive, indicating that a greater amount of learning also increases forecasting error; as with the learning rates, the error is greatest for the lower-level models and least for the two-level models. These observations are not unexpected; it tends to be more difficult to predict the response variable based on a learning curve function if the learning occurs faster and in a larger amount.

Finally, fifteen of the eighteen coefficients for the forgetting rate factors are significant at the 0.05 level, and all but one of these are negative. Since more forgetting results in less learning, it is reasonable to expect that more forgetting results in more accurate forecasts for the same reasons that less and slower learning was seen to result in greater accuracy. As with learning, the impact is greatest for the lower-level models than for the two-level models.

4.3. Pairwise comparisons of different forecasting models

Consider now a different, and perhaps more informative, approach to comparing the relative impact on forecasting error of using a single-level model to estimate future production times when the actual

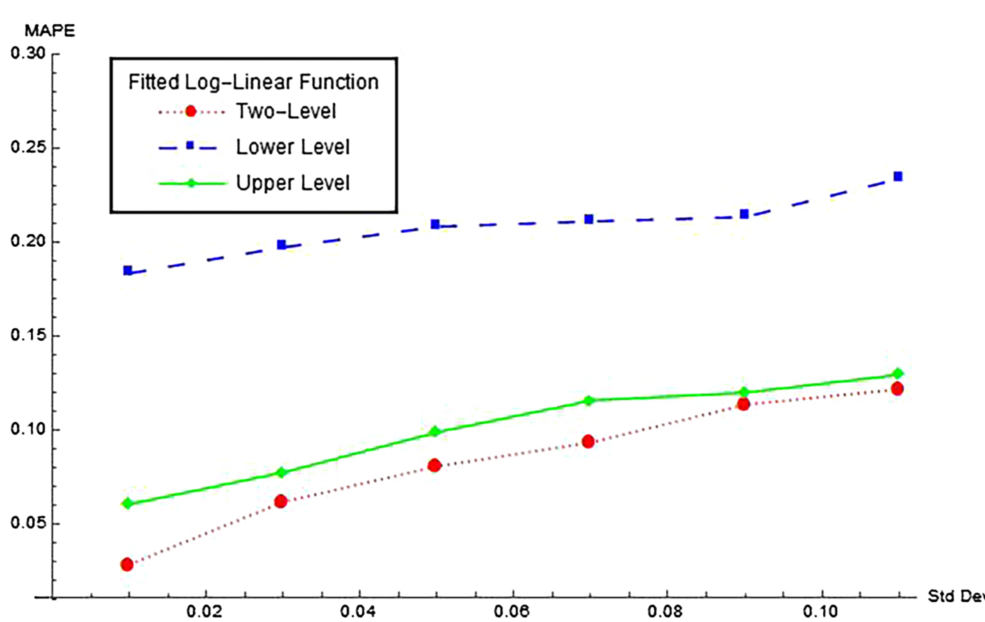

Fig. 2. MAPE for log-linear function for various levels of σ.

Table 4
Comparison of mean MAPE and regression adjusted R-squared for original and high sigma experiments.

		High sigma experiments		Original experiments (Tables 2 & 3)	
Learning function	Model	Mean MAPE	Adjusted R-squared	Mean MAPE	Adjusted R-squared
Log-linear	Two-level	0.1107	0.5323	0.0355	0.8711
Log-linear	Upper-level only	0.1106	0.3005	0.0452	0.6118
Log-linear	Lower-level only	0.1640	0.7574	0.0691	0.9117
Accumulation	Two-Level	0.1073	0.3296	0.0361	0.7815
Accumulation	Upper-level only	0.1075	0.3274	0.0435	0.6631
Accumulation	Lower-level only	0.1147	0.5040	0.0715	0.8975
Replacement	Two-level	0.1260	0.3227	0.0487	0.5431
Replacement	Upper-level only	0.1356	0.3629	0.0641	0.6722
Replacement	Lower-level only	0.1442	0.4600	0.1210	0.7484

Table 5
Pairwise comparisons of forecasting model impacts on MAPE.

		Effect estimates					
		Log-linear function		Accumulation function		Replacement function	
Factor	Factor notation	Upper-two	Lower-two	Upper-two	Lower-two	Upper-two	Lower-two
Initial # of units per production line	n	*−0.0033*	*−0.0156*	*−0.0042*	*−0.0158*	*−0.0095*	*−0.0334*
Standard deviation of error term	σ	*−0.0081*	*−0.0100*	*−0.0052*	*−0.0127*	*−0.0110*	*−0.0339*
Upper-level learning rate	β_1	*0.0032*	*0.0186*	*0.0054*	*0.0066*	*0.0120*	*0.0251*
Lower-level learning rate	β_2	0.0006	*0.0154*	0.0007	*0.0213*	*0.0091*	*0.0562*
Upper-level learning amount	ν_1			0.0010	*−0.0047*	*0.0096*	*0.0246*
Lower-level learning amount	ν_2			−0.0011	*−0.0106*	*0.0056*	*0.0337*
Upper-level forgetting rate	δ_1	−0.0004	*−0.0075*	*−0.0015*	*0.0117*	*−0.0100*	*−0.0113*
Lower-level forgetting rate	δ_2	0.0020	*−0.0073*	−0.0003	*0.0168*	−0.0002	*−0.0209*
Adjusted R-squared		0.2715	0.8234	0.3180	0.8316	0.3574	0.6449

Italicized numbers are significant at the 0.05 level.

learning and forgetting occurs on two levels. For each dataset, let the response variable be the difference in MAPE between either the upper-level or lower-level model and the two-level model; the response variables are then positive if the single-level model performed worse than the two-level model, and negative if the reverse is the case. These differences are then regressed on the same experimental factors as discussed in Section 4.2. The results are shown in Table 5, where all coefficients significant at the 0.05 level are italicized. The coefficients now represent how much the difference in accuracy between the single-level and two-level models increases or decreases as the factor level changes. For the log-linear function, for example, setting n to its high value leads to a decrease in the difference in MAPE between the upper-level and two-level models of 0.0033.

As shown in the table, the coefficients of the first factor n are all negative and significant. As noted above, as n increases, the forecasts become more accurate and the MAPEs of all three models decrease. The benefit of a larger n, however, is more apparent with the single-level models than with the two-level model. As a result, the difference in MAPE between either of the single-level models and the two-level model decreases as we use more units for fitting the models. The coefficients of the second factor σ are also all negative and significant. When the standard deviation of errors is small, the two-level model outperforms both single-level models. However, when the standard deviation becomes bigger, the advantage of the two-level model diminishes. This may be partly due to the large number of parameters in the two-level model; it may not be easy to estimate all parameters in the two-level model with reasonable accuracy when σ is large. The single-level models also suffer with large σ, but they do not seem to be affected as much as the two-level model because the single-level models have only half the number of the parameters of the two-level model. When both n and σ are large, the two-level model could be equaled or even outperformed by the upper single-level model. These observations could explain the satisfactory performance of the single-level models, especially the upper single-level model, in previous studies.

The impact of the six remaining factors is less readily understood. There is a clear pattern for the upper and lower-level learning rates; the coefficients are significant at the 0.05 level in eleven out of twelve cases, and in those eleven cases the coefficients are positive. This evidence indicates that higher learning rates worsen the error gap between the single-level and the two-level models. No equally clear pattern emerges with the eight estimates of the impact of learning amounts, of which six are significant at the 0.05 level. Similarly, only eight of the twelve estimates for the forgetting rate are significant at the 0.05 level, and seven of those are negative. Thus, while the learning amount and forgetting rate results are not as consistently signed as the learning rate estimates, the above results indicate a tendency for the differences between the single-level and the two-level models to get worse as the learning becomes faster and the forgetting becomes slower.

4.4. Performance of different models when no learning and/or forgetting exists

While the above analysis outlines the relative performance of the different models for a range of learning and forgetting parameters, it does not examine the performance of the models when there is no forgetting and/or no learning occurring at one or more of the levels. To consider these possibilities, we conducted fifteen additional experiments, focusing on scenarios where learning and/or forgetting occurred at only one level or (in the case of forgetting) at neither organizational level.

The three basic types of additional "zero forgetting" experiments set either the lower level, upper level, or both the lower and upper level forgetting to zero; for each type of forgetting experiment, all other factors were set to the high or low levels specified in the original experiments to ensure a two-level full factorial design. The two basic types of additional "zero learning" experiments set the lower level

Table 6
Mean absolute percentage errors for scenarios with no learning and/or forgetting.

	Learning function								
	Log-linear function			Accumulation function			Replacement function		
Model	Two-level	Upper-level	Lower-level	Two-level	Upper-level	Lower-level	Two-level	Upper-level	Lower-level
Two-level learning & forgetting	0.0355	0.0452	0.0691	0.0361	0.0435	0.0715	0.0487	0.0641	0.1210
No lower-level forgetting	0.0709	0.0859	0.1738	0.0520	0.0650	0.1109	0.0640	0.1049	0.1583
No upper-level forgetting	0.0488	0.0657	0.1715	0.0394	0.0451	0.1435	0.0592	0.0733	0.2153
No lower- or upper-level forgetting	0.0713	0.0872	0.2057	0.0530	0.0698	0.1661	0.0652	0.1133	0.2346
No lower-level learning or forgetting	0.0292	0.0293	0.0422	0.0322	0.0305	0.0453	0.0337	0.0326	0.0561
No upper-level learning or forgetting	0.0340	0.0452	0.0337	0.0313	0.0412	0.0290	0.0334	0.0416	0.0320

learning and forgetting to zero, or the upper level learning and forgetting to zero; the remaining factors were set, as with the "zero forgetting" experiments, at their original high and low levels to yield a two-level full factorial design. For each of these five types of experiments, three experiments were conducted: one for log-linear functions, one for accumulation functions, and one for replacement functions.

The mean MAPEs for each experiment, along with the mean MAPEs for the original two-level learning and forgetting models shown in Table 2, are displayed in Table 6. Several interesting, but perhaps not unexpected, patterns can be seen from these results. First, all models – two-level, lower-level, and upper-level – perform worse when applied to data where there is learning at two levels but no forgetting at one or more levels. More surprising, perhaps, all three models perform better when applied to scenarios where there is no learning or forgetting at one level. And similar to the original scenario, where learning and forgetting were occurring at both levels, the two-level model also generally outperforms the two single-level models for the three forgetting scenarios; it is matched or slightly outperformed by the upper-level model or the lower-level model only when the single-level models match the actual learning and forgetting scenario, i.e., the upper-level model does best when there is only learning and forgetting at the upper level, and the lower-level model outperforms when there is only learning and forgetting at the lower level.

4.5. Performance of different models when additional explanatory variables are present

So far, we have not considered any explanatory variables other than the variables directly affecting the stock of knowledge. In some situations, additional explanatory variables such as capital, labor or technology-related changes may be known to also impact the performance measure. If the functional relationship of these explanatory variables to the performance measure can be approximated, these variables can be incorporated into the general multi-level learning and forgetting model shown in (1). For purposes of illustration, we will show how a single explanatory variable could be added to the model, and how it's existence could affect the relative performance of two- and single-level models.

Let Z be an additional explanatory variable that increases proportionately to the passage of time. While virtually all levels of the organization can have additional explanatory variables, we consider just a single variable and further assume that the variable is measured when production is started for each item. That is,

$$Z_{[k_M]} = a + b\sum_{i=2}^{k_1} S_{[i]} \tag{6}$$

where a is the initial level of Z at the start time of the first unit, and b is the increase in Z for each unit of time.

With the new variable, the response or the performance measure, Q, is determined in a multiplicative form as follows:

$$Q_{[k_M]} = \alpha\prod_{m=1}^{M} F^m(E^m_{[k_m]})G(Z_{[k_M]})e^{\varepsilon_{[k_M]}},\ \ \varepsilon_{[k_M]} \sim Normal\ (0, \sigma). \tag{7}$$

Among the three learning curve functions, since the log-linear function is the most widely used functional form in the literature, we consider the log-linear function for purposes of illustration. We assume that the functional relationship between the response variable and the new variable also follows a log-linear form:

$$G(Z_{[k_M]}) = Z_{[k_M]}^{-\gamma}, \tag{8}$$

where γ is a parameter representing how intensely Z affects the response. We vary γ from 0 to 2.0 with an increment of 0.25 to model situations where the impact of Z on Q decreases ($0 < \gamma < 1$), remains constant ($\gamma = 1$), or increases ($1 < \gamma \leqslant 2$) over time. Note that when $\gamma = 0$, the new variable has no impact on the response, i.e., Eq. (7) reduces to the original mathematical representation given in Eq. (1). All other parameters are set so that the advantage of the two-level model over the single-level models is maximized. Specifically, we consider that case when the data exhibit a small amount of dispersion ($\sigma = 0.01$), the number of units used for fitting is small ($n = 5$), the learning rates are high ($\beta_1 = 0.211,\ \beta_2 = 0.354$), and the forgetting rates are low ($\delta_1 = \delta_2 = 0.1$). To generate values of $G(Z_{[k_M]})$ that are not grossly out of scale with the $F^m(E^m_{[k_m]})$ for the extreme values of γ, we set $a = 1.0$ and $b = 0.25$, yielding Z equal to one for the first item being produced and increases at the rate of 0.25 per unit of time thereafter. When fitting the single-level models to the data, we include $G(Z_{[k_M]})$ in the model being fitted along with the single-level component.

Our results are shown in Fig. 3. Each point plotted on the chart is the mean of thirty observations. Note that the two-level model consistently outperforms the two single-level models, with the upper-level model generally outperforming the lower-level model. The performance of all three models deteriorates as γ increases; if the new variable has an increasing impact on production over time, mis-specifying the model leads to increasing forecasting errors. Furthermore, the degree of deterioration is quite significant with the lower-level model but is relatively small with the two-level model. The performance of the upper-level model falls in the middle.

If we reverse the experimental parameters to lessen the advantage of the two-level model ($\sigma = 0.05,\ n = 15,\ \beta_1 = 0.121,\ \beta_2 = 0.204,\ \delta_1 = 0.2,\ \delta_1 = 0.2$), Fig. 4 shows that for values of γ below one, the performance of the single- and two-level models are quite similar. As γ increases beyond one, however, the performance of all three models deteriorates somewhat, with the lower-level model becoming significantly worse. Unlike the previous case when the advantage of the two-level model over the single-level models is maximized, the performance of the two-level model is not distinctively different from that of the upper-level model.

In situations where the additional explanatory variable Z becomes increasingly important in estimating Q as time passes, therefore, the performance of all three models declines, although the two-level model continues to dominate, followed by the upper-level and then the lower-level models. Furthermore, the same parameter combination that was

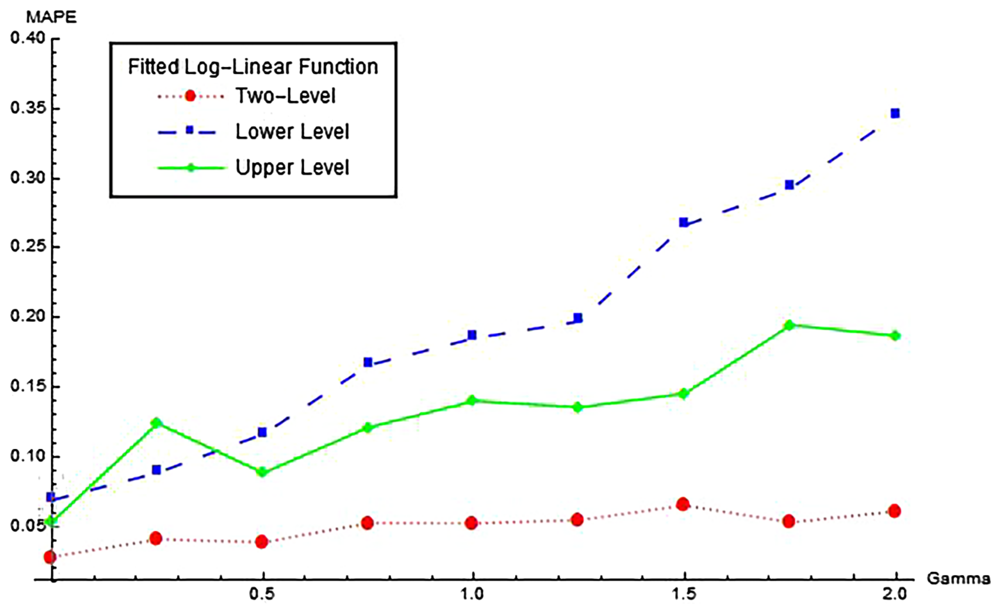

Fig. 3. MAPE for high-learning log-linear function with added variable *Z*.

found to favor the two-level model in the absence of additional explanatory variables exacerbates the relative differences between the models when a new explanatory variable is present.

4.6. Summary of model performance

The experimental results when learning and forgetting occurs at both levels can be summarized as follows:

- In terms of average forecasting accuracy the upper single-level model performs nearly as well as the two-level model, and it could equal or even outperform the two-level model when n and σ are large.
- As the standard deviation of errors σ increases, the MAPE increases in both the two-level and the single-level models, and it increases faster with the two-level model than with the single-level models.
- As the number of ships used for fitting n increases, the MAPE decreases for all models, and the gains are greatest for the single level models.
- As the rate or amount of learning increases, the MAPE increases. It would be more difficult to predict the response variable based on a learning curve as the learning occurs faster or in a larger amount. The error gap between the single-level and two-level models expands as the learning rates increases. However, no clear pattern is observed between the error gap and the amount of learning.
- For the same reason as the learning rate, as the forgetting rate increases, the MAPE decreases. The error gap between the single-level and two-level models mostly shrinks as the forgetting rates increase.
- The upper single-level model performs better than the lower single-level model in virtually all cases. Especially when n is small, its superiority over the lower single-level model stands out more distinctively.

When learning and forgetting do not both occur at two levels, we

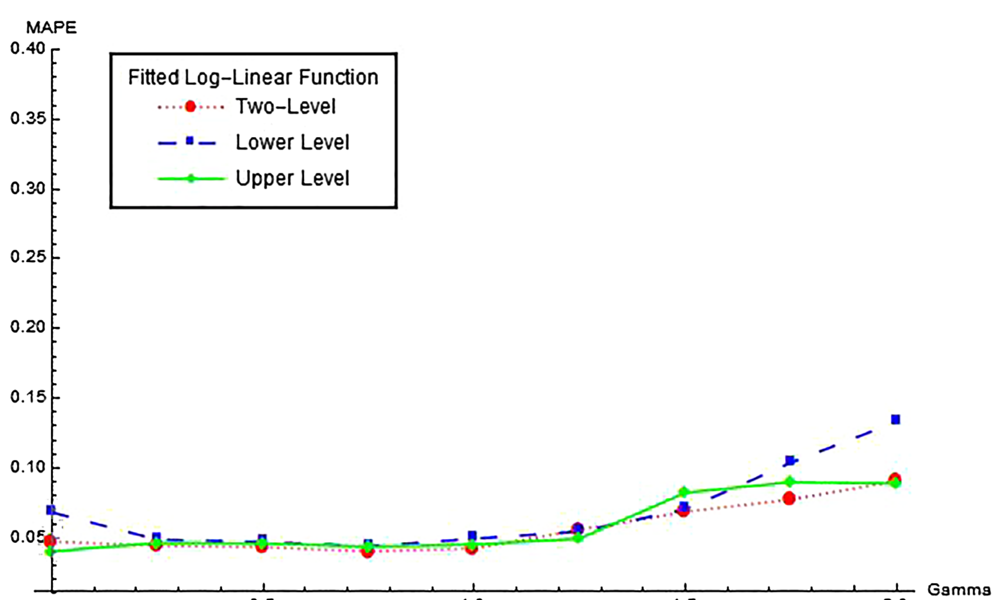

Fig. 4. MAPE for low-learning log-linear function with added variable *Z*.

observe that:

- The two-level model generally outperforms both single level models when there is learning at both levels but no forgetting at one or more levels.
- The two-level model performs almost as well as the correctly specified single-level model when there is no learning or forgetting at one level.

As discussed in the introduction, the learning curve has played an important role in the estimation of future production time, the accuracy of which is critical for decisions regarding scheduling, inventory management and workforce assignment. In many practical situations, due to a lack of sufficiently detailed data, it is often necessary to rely on the conventional aggregated single-level model for these production time estimates. For managers interested primarily in accurate forecasts of future production times, rather than precise estimates of the extent of learning and forgetting at different levels in the organization, our findings provide a practical guidance on when the upper-level approximation works well and how much uncertainty the approximation generates. More broadly, the solid performance, under a wide range of conditions, of the upper-level approximation in estimating production times generated under conditions of two-level learning confirms the utility of the commonly used single-level learning models.

More specifically, our results show that the upper single-level model approximates the total learning and forgetting occurring at two levels generally well unless the volatility of the performance measure is very small. In the Liberty ships dataset, the estimated standard deviation of the error term was about 0.1 (since we assume a multiplicative model as shown in Eq. (1), this is equivalent to $e^{0.1} = 1.105$ or approximately 10.5% forecasting errors on average), at which the upper single-level model performed nearly equally as the two-level model. Caution in using the upper single-level model is warranted, however, in the early stages of a learning curve when n is small or the learning curve is steep; in these situations the upper single-level approximation did not perform well and a more careful monitoring of learning at various levels of an organization is suggested. After the early stages of learning have passed (in our experiments when $n = 10$ or higher), we can safely apply the upper single-level approximation to predict the performance measures.

5. Conclusions

Learning can occur at every level of an organization, and the rates of learning can vary from one level to another. Most conventional learning models measure learning and forgetting at a single aggregated level and implicitly assume that this is a good approximation of the total learning and forgetting for the entire organization. Recently, Park et al. (2013) proposed a two-level learning model, which is capable of measuring learning and forgetting at two different levels of an organization. While a multiple-level model may be closer to the real process of learning in an organization, due to the lack of disaggregated data it may still be necessary to use a conventional single-level model for prediction. This paper compared the performance of the single- and two-level models using extensive numerical experiments based upon the Liberty ships dataset. Datasets were generated according to the two-level model, and two versions of the single-level model as well as the two-level model were then fitted to the datasets. This enabled the authors to determine how accurately the aggregated single-level models could estimate the learning and forgetting which actually occurred at two levels of an organization.

Single-level models may not be appropriate for accurately measuring the rates of learning and forgetting at each level, but single-level models may provide quite accurate estimates for the performance measure of the entire organization; this is most likely in scenarios where a large degree of randomness is present in the data or where future learning is expected to be relatively modest. This last situation can occur not only when learning rates are low and/or forgetting rates are high, but also when the cumulative number of items produced has moved the organization to the relatively flatter part of the learning curves. While organizations may be unwilling to delay model fitting to improve the accuracy of a single-level model, the wide range of learning rates and forgetting rates reported in the literature for different industries, as well as the noisiness expected in field data, suggests that there are likely many situations in which one or more of the above criteria are satisfied.

As discussed early in the paper, the Liberty ships dataset was used to set the parameters of our numerical experiments. While we feel that the parameter ranges set in this manner were representative of a wide range of applications, our results are limited to these ranges, and different parameter ranges could be used to further test our conclusions. Furthermore, other learning and forgetting functions – such as the power functions used by Sikstrom and Jaber (2012) – could be used to test the robustness of our results. A more direct test, perhaps, would be to apply the approaches discussed in this paper to other real datasets where the learning and forgetting processes differ from the Liberty ships case.

In addition, while we briefly examined the impact of additional explanatory variables on the relative performance of the single- and two-level models, these results are based on just two somewhat extreme cases where the advantage of the two-level model is maximized or minimized; a more thorough numerical analysis would be required before we could make a firm conclusion on the impact of additional explanatory variables. Another possible extension of this study would be to examine the performance of all three models, especially the accuracy of the single-level approximation, in the presence of various types of additional explanatory variables.

More fundamentally, the current model assumes that the learning and forgetting rates at the shipway level are identical across all shipways. This assumption may be valid in many organizations, especially manufacturing plants, which consist of multiple and nearly-homogeneous entities in terms of the types of equipment, the scope of procedures, and the skills of workforce. In organizations without such homogeneity, however, use of the proposed two-level model without further modification would not be valid. One possible extension of the two-level model, therefore, is to consider an organization consisting of heterogeneous lower-level entities, representing a situation where entities using quite different processes and producing different products are managed by the same upper-level organization. In such a scenario, each entity at the lower level is allowed to have its own learning and forgetting rates. The heterogeneity in the lower level is most likely to increase the variability of the performance measure or the response variable. As we have seen with σ, the standard deviation of the error term, increased variability disadvantages the two-level model, and this same effect could result from an increase in lower-level heterogeneity. On the other hand, different learning and forgetting rates could make the aggregation by a single-level model more difficult, and result in less accurate forecast for the upper single-level model. Further experimentation will be necessary to examine how the heterogeneity affects the accuracy of the approximation by a single-level model.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Argote, L. (1999). *Organizational learning: Creating, retaining and transferring knowledge.* New York, New York: Springer.

Argote, L., Beckman, S., & Epple, D. (1990). The persistence and transfer of learning in industrial settings. *Management Science, 36*(2), 140–154.

Bailey, C. (1989). Forgetting and the learning curve: A laboratory study. *Management Science. 35*(3), 340–352.

Balasubramanian, N., & Lieberman, M. (2011). Learning-by-doing and market structure. *Journal of Industrial Economics, 59*(2), 177–198.

Benkard, C. L. (2000). Learning and forgetting: The dynamics of aircraft production. *American Economic Review, 90*(4), 1034–1054.

Box, G. E. P., Huner, W. G., & Hunter, J. S. (1978). *Statistics for Experimenters.* New York: Wiley New York.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *American Economic Review, 18*(Supplement), 139–165.

Darr, E., Argote, L., & Epple, D. (1995). The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Science, 41*(11), 1750–1762.

Epple, D., Argote, L., & Murphy, K. (1996). An empirical investigation of the microstructure of knowledge acquisition and transfer through learning by doing. *Operations Research, 44*(1), 77–86.

Glock, C. H., & Jaber, M. Y. (2014). A group learning curve model with and without worker turnover. *Journal of Medelling in Management. 9*(2), 179–199.

Grosse, E., Glock, C., & Christoph, H. (2015b). Decision support models for production ramp-up: a systematic literature review. *International Journal of Production Research, 53*, 6637–6651.

Grosse, E., Glock, C., & Christoph, H. (2015a). The effect of worker learning on manual order picking processes. *International Journal of Production Economics, 170*, 882–890.

Grosse, E., Glock, C., Christoph, H., & Müller, S. (2015). Production economics and the learning curve: A meta-analysis. *International Journal of Production Economics, 170*, 401–412.

Harvey, R. A. (1979). Learning in production. *The Statistician, 28*(1), 39–57.

Heimerl, C., & Kolisch, R. (2010). Work assignment and qualification of multi-skilled human resources under knowledge depreciation and company skill level targets. *International Journal of Production Research, 48*(13), 3759–3781.

Jaber, M. Y. (2006). Learning and forgetting models and their applications. In A. B. Badiru (Ed.). *Handbook of industrial and systems engineering* (pp. 1–27). Baco Raton, FL: CRC Press.

Jaber, M. Y. (Ed.). (2011). *Learning curves: Theory, models and applications*. Baco Raton, FL: CRC Press.

Kim, I., Moskowitz, H., Plante, R. D., Seo, H., & Tang, K. (2007). Organizational learning with continuous depreciation of knowledge. *International Journal of Operations and Quantitative Management, 13*(2), 77–94.

Kim, I., & Seo, H. (2009). Depreciation and transfer of knowledge: An empirical exploration of a shipbuilding process. *International Journal of Production Research, 47*(7), 1657–1876.

Nadeau, M., Kar, A., Roth, R., & Kirchain, R. (2010). A dynamic process-based cost modeling approach to understand learning effects in manufacturing. *International Journal of Production Economics, 128*(1), 223–234.

Nembhard, D. A., & Bentefouet, F. (2015). Selection, grouping, and assignment policies with learning-by-doing and knowledge transfer. *Computers and Industrial Engineering, 79*, 175–187.

Nembhard, D. A., & Uzumeri, M. V. (2000). An individual-based description of learning within an organization. *IEEE Transactions on Engineering Management, 47*(3), 370–378.

Park, Y., Springer, M., & Kim, I. (2013). Organizational learning and forgetting: A two-level approach. *International Journal of Operations and Quantitative Management, 19*(4), 275–293.

Restle, F., & Greeno, J. G. (1970). *Introduction to mathematical psychology*. New York, New York: Addison-Wesley.

Sikstrom, S., & Jaber, M. Y. (2012). The depletion-power-integration-latency (DPIL) model of spaced and massed repetition. *Computers and Industrial Engineering, 63*, 323–337.

Suer, G. A., & Tummaluri, R. R. (2008). Multi-period operator assignment considering skills, learning and forgetting in labour-intensive cells. *International Journal of Production Research, 46*(2), 469–493.

Thompson, P. (2007). How much did the Liberty shipbuilders forget? *Management Science, 53*(6), 908–918.

Van Peteghem, V., & Van Houcke, M. (2015). Influence of learning in resource-constrained project scheduling. *Computers and Industrial Engineering, 87*, 569–579.

Wright, T. (1936). Factors affecting the cost of airplanes. *Journal of Aeronautical Science. 2*, 122–128.

Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision Sciences, 10*, 302–328.